AD-A102 058 MARYLAND UNIV COLLEGE PARK COMPUTER VISION LAB F/G 12/1
GENERALIZED BLOMQVIST CORRELATION.(U)
APR 81 S M DUNN AFOSR-77-3271
UNCLASSIFIED TR-1043 AFOSR-TR-81-0560 NL

1 of 1
AD
A102058

END
DATE
FILMED
8 81
DTIC

AFOSR-TR- 81 -0560 LEVEL II

10

AD A102058

COMPUTER SCIENCE
TECHNICAL REPORT SERIES

UNIVERSITY OF MARYLAND
COLLEGE PARK, MARYLAND
20742

DTIC
ELECTE
JUL 28 1981
S D

D

Approved for public release;
distribution unlimited.

DTIC FILE COPY

81 7 24 049

LEVEL II

10

TR-81-0569

TR-1043
AFOSR-77-3271

April 1981

# GENERALIZED BLOMQVIST CORRELATION

Stanley M. Dunn

Computer Vision Laboratory
Computer Science Center
University of Maryland
College Park, MD 20742

## ABSTRACT

A distribution-free measure of dependence between two random variables is defined and shown to be a generalization of the quadrant measure of association. This new measure is presented as a U-statistic and its asymptotic efficiency relative to the quadrant measure is given. This measure can be used for locating edges in computer images.



The support of the U.S. Air Force Office of Scientific Research under Grant AFOSR-77-3271 is gratefully acknowledged, as is the help of Janet Salzman in preparing this paper. The author would like to thank Drs. A. Rosenfeld and P. Thrift for their assistance, and Mr. R. Chellappa whose guidance was invaluable.

## 1. Introduction

The problem of detecting edges in computer image processing can be formulated as a problem of determining the correlation between an $n \times n$ neighborhood of the image and a template containing an ideal edge. This is a special case of the more general problem of measuring the dependence between two random variables. The Pearson product-moment (PPM) correlation is the most commonly used measure, yet there are underlying assumptions that are often neglected when PPM correlation is applied in practice.

Statisticians have developed many methods for the measurement of dependence that do not have the same restrictions that PPM does. A good introduction to several of these is Kruskal [1]. Included in this survey is a discussion of the quadrant measure of association. This measure is also presented in Blomqvist [2] where it is referred to as the double median test for association.

In Blomqvist [2] as well as Elandt [3], the double median test is called a non-parametric test of tendency. Here we call the double median test and its generalization distribution-free tests. We have adopted the following convention as presented in Gibbons [4]:

Definition 1: A distribution free method is one based on functions of sample observations whose corresponding random variables have distributions which do not depend on the specific distribution function of the population from which the sample was drawn.

Definition 2: A nonparametric test is a test for a hypothesis which is not a statement about parameter values.

Definition 3: A parameter is a characteristic of the population. A parameter can be thought of as an unspecified constant appearing in a family of probability distributions, but it can also be thought of as any characteristic of the population, properly including those constants appearing in probability distributions.

In this paper we adopt the broader definition of parameter. In this sense, a median or other order statistic is a population parameter, for example. Typically, in using distribution-free methods, the underlying assumption is that the population is continuous. We make that assumption here.

We now discuss Blomqvist's double median test for association, followed by its generalization, referred to as generalized Blomqvist correlation (GBC). Some of the properties of GBC are discussed, and it is snown that GBC is assymptotically more efficient than Blomqvist's double median test.

## 2. Blomqvist's double median test

As discussed in Section 1, the motivation behind distribution-free methods is for these methods to be valid under weak assumptions about the population distribution. In addition to this, Blomqvist [2] also gave as a motivation the notion that such tests should be easy to deal with in practice. This aspect will be given further attention later in this paper. The assumption made about the population is that the cdf $F(x,y)$ is assumed to have continuous marginal cdfs $F_1(x)$ and $F_2(y)$. This is done so that $\mathrm{Prob}\{x_i=x_j\}$ or $\mathrm{Prob}\{y_i=y_j\}=0$ for $1\le i\le n$ and $1\le j\le n$ and $i\ne j$.

Let $(x_1,y_1),\ldots,(x_n,y_n)$ be a sample with such a cdf $F(x,y)$. Denote by $\pi_s$ the probability

$$\pi_s=\mathrm{Prob}\{(x<x_0 \text{ and } y<y_0) \text{ or } (x>x_0 \text{ and } y>y_0)\} \quad (1)$$

for some $x_0$ and $y_0$. Similarly denote by $\pi_d$ the probability

$$\pi_d=\mathrm{Prob}\{(x<x_0 \text{ and } y>y_0) \text{ or } (x>x_0 \text{ and } y<y_0) \quad (2)$$

for the same $x_0$ and $y_0$. The above equations can be rewritten as

$$\pi_s=\mathrm{Prob}\{(x-x_0)(y-y_0)>0\}$$
$$\pi_d=\mathrm{Prob}\{(x-x_0)(y-y_0)<0\} \quad (3)$$

Here one can easily see that $\pi_s$ is the probability that the deviations of $x$ and $y$ from $x_0$ and $y_0$ have the same sign, whereas $\pi_d$ is the probability of different signs.

We define as a measure of correlation the difference in these two probabilities. As in Kruskal [1], we denote this measure by

$$\gamma=\pi_s-\pi_d$$

Although the choice of $x_0$ and $y_0$ is arbitrary, we choose to let $x_0=x_m$, the median of x, and let $y_0=y_m$, the median of y. $\gamma$ will be 1 iff x and y deviate from their medians in the same direction, and -1 iff x and y deviate from their medians in opposite directions. Clearly if x and y are independent $\pi_s=\pi_d$ and so $\gamma$ is zero.

We construct the sample analog in the following way. The xy plane is divided into four regions by the lines $x=x_m$ and $y=y_m$ (see Figure 1). The four quadrants will be labelled by the Roman numerals I, II, III, IV as follows:

Quadrant I: Those pairs $(x_i,y_i)$ such that $x_i<x_m$ and $y_i<y_m$.

Quadrant II: Those pairs $(x_i,y_i)$ such that $x_i>x_m$ and $y_i<y_m$.

Quadrant III: Those pairs $(x_i,y_i)$ such that $x_i>x_m$ and $y_i>y_m$.

Quadrant IV: Those pairs $(x_i,y_i)$ such that $x_i<x_m$ and $y_i>y_m$.

From the above definitions, it is clear that the number n of samples used is even. Should the sample size be odd, then either one or two samples fall exactly on the lines $x=x_m$ and $y=y_m$. If there is one point (the point $(x_m,y_m)$) we do not count this. If there are two points (the points $(x_m,y_i)$ and $(x_j,y_m)$ for some i and j), we do not count one of them and we let the other one be included in both regions it touches. If the sample size is even, there is no problem and the medians are defined by

$$x_m=(x_{\frac{n}{2}}+x_{\frac{n}{2}+1})/2$$

$$y_m=(y_{\frac{n}{2}}+y_{\frac{n}{2}+1})/2$$

Let $n_1$ be the number of points in quadrants I and III. Similarly, let $n_2$ be the number of points in quadrants II and IV. Clearly,

$n=n_1+n_2$. We define as a measure of dependence between the two random variables x and y the statistic q' where

$$q'=\frac{n_1-n_2}{n_1+n_2} \qquad (4)$$

The statistic q' lies between -1 and 1, which is desirable for a measure of dependence.

One of the earliest presentations of this statistic was in Mosteller [5]. This presentation used two lines to partition the xy plane in the x direction, and one in the y direction. The double median test is the limiting case when these two lines coincide at the median. An alternative definition of q' in terms of U-statistics was given in Elandt [3], and this is the method used in the next section. The estimate q' was given in Blomqvist [2] along with its asymptotic distribution and asymptotic relative efficiency (ARE). It was shown that the ARE of q' relative to r (the estimate of the PPM correlation) is about 41%.

The interested reader is also directed to Bradley [6] for an elementary discussion of this test and others based on order statistics.

## 3. A generalization of the double median test

The test statistic q' is computed by counting the number of sample points in the four quadrants of the xy plane. The lines which divide the xy plane are $x=x_m$ and $y=y_m$, where $x_m$ and $y_m$ are the x and y sample medians, respectively. Here we investigate what happens when additional order statistics are introduced to divide the xy plane into smaller regions.

Before proceeding it is necessary to introduce some notation. Let N be the number of sample points, where N is assumed to be even. If N is odd, it is made even by a procedure similar to that described in Section 2 for the double median test. Let n be the number of regions into which the x and y axes have been divided. Hence the total number of regions in the xy plane is $n^2$.

We divide the x and y axes into n regions by introducing n-1 order statistics from the x and y samples, respectively. The correlation scheme that uses n-1 order statistics is called "Generalized Blomqvist Correlation of Order n."

Let $\xi_i^{(n)}$ denote the ith x order statistic from GBC of order n. Similarly, let $\eta_i^{(n)}$ denote the ith order statistic from the same correlation scheme. Using this notation one can see that

$$\xi_1^{(2)}=x_m$$

$$\eta_1^{(2)}=y_m$$

Analogous to the definitions in Section 2, we denote by $\pi_s^{(n)}$ the probability

$$\pi_s^{(n)}=\text{Prob}\{(\xi_{i-1}^{(n)}<x<\xi_i^{(n)} \text{ and } (\eta_{i-1}^{(n)}<y<\eta_i^{(n)})\} \qquad (5)$$

for $1 \le i \le n$, and by $\pi_d^{(n)}$ the probability

$$\pi_d^{(n)} = \mathrm{Prob}\{(\xi_{i-1}^{(n)} < x < \xi_i^{(n)}) \text{ and } (\eta_{n-i}^{(n)} < y < \eta_{n-i+1}^{(n)})\} \qquad (6)$$

for $1 \le i \le n$. $\xi_0^{(n)}$ and $\eta_0^{(n)}$ are taken to be $-\infty$ and $\xi_n^{(n)}$ and $\eta_n^{(n)}$ are taken to be $+\infty$.

Note that in the double median test $\pi_s + \pi_d = 1$. Here $\pi_s^{(n)} + \pi_d^{(n)} = 1$ iff $n=2$. Hence we define

$$\pi_r^{(n)} = \mathrm{Prob}\{x \notin (\xi_{i-1}^{(n)}, \xi_i^{(n)}) \text{ or } y \notin (\eta_{i-1}^{(n)}, \eta_i^{(n)}) \text{ or } y \in (\eta_{n-i}^{(n)}, \eta_{n-i+1}^{(n)})\} \quad (7)$$

$\pi_s^{(n)}$ is the probability that both x and y lie in the interval bounded by their order statistics of the same rank. $\pi_d^{(n)}$ is the probability that y lies in the interval bounded by order statistics whose ranks are related to the ranks of the x order statistics by the equation $j=n-i+1$. Here j is the rank of y order statistic and i is the rank of x order statistic. $\pi_r^{(n)}$ is the probability that x and y lie in other regions of the xy plane, e.g., $\pi_r^{(n)} = 1-(\pi_s^{(n)} + \pi_d^{(n)})$.

We define as a measure of association $\gamma_{(n)}$, where $\gamma_{(n)}$ is given by

$$\gamma_{(n)} = \pi_s^{(n)} - \pi_d^{(n)} \qquad (8)$$

By considering the xy plane as an $n \times n$ matrix $\underset{\sim}{R}$, the probability $\pi_s^{(n)}$ is the probability that the sample $(x_i, y_i)$ lie in the regions along the main diagonal. $\pi_d^{(n)}$ is the probability that the sample $(x_i, y_i)$ lie in the regions along the diagonal from the lower left to the upper right corner. $\pi_r^{(n)}$ is the probability that $(x_i, y_i)$ falls in the other regions of the xy plane. If $n=2$, $\pi_r^{(2)} = 0$ since there are no other regions.

The sample statistic $q'_{(n)}$ is computed in the following manner. Denote by $r_{ij}^{(n)}$ the region of the xy plane such that

$$\xi_{i-1}^{(n)} < x < \xi_i^{(n)} \quad \text{and} \quad \eta_{j-1}^{(n)} < y < \eta_j^{(n)} \tag{9}$$

Let $|r_{ij}^{(n)}|$ denote the number of samples in the region $r_{ij}^{(n)}$. The sample statistic $q'_{(n)}$ is computed by

$$q'_{(n)} = \frac{1}{N}\left(\sum_{i=1}^{n} |r_{ii}^{(n)}| - \sum_{i=1}^{n} |r_{i,n-i+1}^{(n)}|\right) \tag{10}$$

Let
$$n_1 = \sum_{i=1}^{n} |r_{ii}^{(n)}|$$
$$n_2 = \sum_{i=1}^{n} |r_{i,n-i+1}^{(n)}|$$

Then
$$q'_{(n)} = \frac{n_1 - n_2}{N} \tag{11}$$

If there is a positive correlation between x and y then $n_1 > n_2$ and $q'_{(n)}$ will approach +1. If there is a negative correlation between x and y, then $n_2 > n_1$ and $q'_{(n)}$ approaches -1. If x and y are independent then $n_1 = n_2$ and $q'_{(n)} = 0$.

We now examine an alternative formulation of the sample statistic $q'_{(n)}$ in terms of U statistics as described in Hoeffding [7]. By showing that $q'_{(n)}$ is a U statistic, we can immediately determine both the asymptotic distribution and the variance of $q'_{(n)}$. From this, the ARE follows immediately.

## 4. The U Statistic Form of $q'_{(n)}$

Hoeffding [6] shows that statistics of the form

$$U = \frac{1}{n(n-1)\ldots(n-m+1)} \Sigma'' \Phi(x_{\alpha_1}, \ldots x_{\alpha_m}) \tag{12}$$

(where the summation is over all permutations $\alpha_1$ to $\alpha_m$ of the integers 1 to n) are unbiased estimators of their population characteristics $\theta$. It is also shown that $\sqrt{n}\,(U-\theta)$ tends to a normal distribution as $n \to \infty$. It is desirable that the statistic $q'_{(n)}$ have these properties, and hence we demonstrate a U statistic formulation of $q'_{(n)}$.

Randles and Wolfe [8] also provide a description of U statistics. They do point out, however, the importance of the estimability of the population parameter. Hence we state and prove

<u>Theorem 1</u>: $\gamma_{(n)}$ is estimable of degree 1.

<u>Proof</u>: To show that $\gamma_{(n)}$ is estimable of degree 1, we show that there exists a function $\phi$ such that

$$E\{\phi(x_1)\} = \gamma_{(n)}$$

Let $Z_i$ be the sample $(x_i, y_i)$ and let $\phi(Z_i)$ be

$$\phi(Z_i) = \sum_{j=1}^{n} \sum_{k=1}^{n} \psi(x_i, y_i, j, k) \tag{13}$$

where $\psi(x_i, y_i, j, k) =$

$$\begin{cases} 1, (j=k) \text{ and } (\xi_{j-1}^{(n)} - x_i)(x_i - \xi_j^{(n)})(\eta_{k-1}^{(n)} - y_i)(y_i - \eta_k^{(n)}) > 0 \\ -1, (k=n-j) \text{ and } (\xi_{j-1}^{(n)} - x_i)(x_i - \xi_j^{(n)})(\eta_{k-1}^{(n)} - y_i)(y_i - \eta_k^{(n)}) > 0 \\ 0, \text{ otherwise} \end{cases} \tag{14}$$

Now we must show that $E\{\phi(Z_i)\} = \gamma_{(n)}$ for all distributions $F$ in the family $\mathcal{F}$. It is clear that $\phi(Z_i)$ does not depend on $F(\cdot)$ in any way. We are considering all distributions $F(\cdot)$ under the null hypothesis $H_0$: $F(x,y)=F_1(x)F_2(y)$, that is, the hypothesis that x and y are independent.

Under this hypothesis, $\gamma_{(n)}=0$, so it suffices to show that $E\{\phi(Z_i)\}=0$. Under this hypothesis, all values of $\psi(x_i,y_i,j,k)$ are equally likely, so that

$$E\{\phi(Z_i)\} = E\{\sum_{j=1}^{n} \sum_{k=1}^{n} \psi(x_i y_i,j,k)\}$$

$$= \sum_{j=1}^{n} \sum_{k=1}^{n} E\{\psi(x_i,y_i,j,k)\}$$

$$= 0$$

which is the desired result: $\gamma_{(n)}$ is estimable of degree 1.

Knowing that $\gamma_{(n)}$ is estimable of degree 1, we can construct the U statistic

$$U = \frac{1}{N} \Sigma'' \phi(Z_i) \qquad (15)$$

where $\phi(Z_i)$ is defined in equation (13), $\psi(x_i,y_i,j,k)$ is defined in equation (14) and the sum $\Sigma''$ extends over all $i$ such that $1 \le i \le N$.

By the results of Hoeffding [7], we know that U as given in equation (15) is an unbiased estimate of $\gamma_{(n)}$. Also $\sqrt{N}\ (U-\gamma_{(n)})$ has a limiting normal distribution as $N \to \infty$.

Now it is of interest to determine the variance of the statistic U. With this in hand, the ARE of U relative to q' can be computed.

The variance of U is computed by the following equation from Randles and Wolfe [8]:

$$\mathrm{Var}(U) = \frac{1}{\binom{N}{r}} \sum_{c=1}^{r} \binom{r}{c}\binom{N-r}{r-c}\zeta_c$$

Since the estimability r is 1 here,

$$\mathrm{Var}(U) = \frac{1}{N}\,\zeta_1 \tag{16}$$

where $\zeta_1$ is given by

$$\zeta_1 = E\{\phi^2(Z_i)\} - \gamma^2_{(n)} \tag{17}$$

Again we compute this expectation under the null hypothesis of independence. The values of $\psi(x_i,y_i,j,k)$ are equally likely and therefore

$$E\{\phi^2(Z_i)\} = E\{(\sum_{j=1}^{n}\sum_{k=1}^{n} (x_i,y_i,j,k))^2\}$$

$$= E\{\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{j'=1}^{n}\sum_{k'=1}^{n} \psi(x_i,y_i,j,k)\psi(x_i,y_i,j',k')\}$$

The product is 1 for j=j' and k=k' and otherwise 0. Since 1 is twice as likely as 0 ($\psi$ takes on -1,0, or 1), the expectation can now be solved:

$$E\{\phi^2(Z_i)\} = E\{\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{j'=1}^{n}\sum_{k'=1}^{n} \psi(x_i,y_i,j,k)\psi(x_iy_ij',k')\}$$

$$= \frac{2}{3}\left(\frac{2n}{n^2}\right) = \frac{4}{3n} \tag{18}$$

Hence equation (17) becomes

$$\zeta_1 = E\{\phi^2(Z_i)\} - \gamma^2_{(n)}$$

$$= \frac{4}{3n} - \gamma^2_{(n)} \tag{19}$$

Put equation (19) into equation (16) to obtain the variance of the statistic U:

$$\begin{aligned} \mathrm{var}(U) &= \frac{1}{N}\,\zeta_1 \\ &= \frac{1}{N}\left(\frac{4}{3n} - \gamma^2_{(n)}\right) \qquad (20) \end{aligned}$$

This result is used in the next section to compute ARE(U,q').

## 5. The Asymptotic Relative Efficiency of U

The asymptotic relative efficiency of two statistical tests is an indication of the relative power of the tests. That is to say, it is a measure of the ratio of the sample sizes the two tests require to achieve the same level of statistical significance. The interested reader is referred to Gibbons [4] for a detailed explanation and several worked out examples.

Here we will investigate the ARE of U relative to q'. We would like to see if ARE(U,q') is at least 1 for all n. The ARE is defined by

$$\mathrm{ARE}(U,q') = \lim_{N \to \infty} \frac{e(U)}{e(q')} \tag{21}$$

where $e(\cdot)$ is the efficacy of the test statistic:

$$e(T) = \frac{[dE(T)/d\theta]^2}{\sigma^2(T)}\bigg|_{\theta=\theta_0} \tag{22}$$

E(T) is the expected value of the test statistic T, and $\theta$ is the population parameter. In both cases $dE(T)/d\theta$ is 1 since $E(q') = \gamma$ and $E(U) = \gamma_{(r)}$. Thus, equation (21) becomes

$$\mathrm{ARE}(U,q') = \lim_{N \to \infty} \frac{\sigma^2(q')}{\sigma^2(U)} \tag{23}$$

It is also necessary to indicate the hypotheses being tested, since in the process of taking the limit we are also letting the alternate hypothesis approach the null hypothesis. As previously indicated in this paper our hypotheses are

$$H_0: \quad \gamma=0$$

$$H_1: \quad \gamma\neq0$$

From Blomqvist [2], the variance of q' is $\frac{4a_0(1-2a_0)}{k}$. From equation (20) the variance of U is $\frac{1}{N}(\frac{4}{3n} - \gamma^2_{(n)})$. Note that the k in the expression of var(q') is equal to $\frac{N}{2}$ from the definitions in Blomqvist [2]. We now compute ARE(U,q'):

$$\begin{aligned} \mathrm{ARE}(U,q') &= \lim_{N \to \infty} \frac{\sigma^2(q')}{\sigma^2(U)} \\ &= \lim_{N \to \infty} \left[\frac{4a_0(1-2a_0)}{k} \Big/ \frac{1}{N}\left(\frac{4}{3n} - \gamma^2_{(n)}\right)\right] \\ &= \lim_{N \to \infty} \left[\frac{4a_0(1-2a_0)}{k} \cdot \frac{3Nn}{4-3n\gamma^2_{(n)}}\right] \end{aligned}$$

Since N=2k,

$$\mathrm{ARE}(U,q') = \lim_{N \to \infty} \left[\frac{8a_0(1-2a_0)3n}{4-3\gamma^2_{(n)}}\right]$$

Now taking the limit, $\gamma^2_{(n)}=0$, $a_0=\frac{1}{4}$

$$\mathrm{ARE}(U,q') = \frac{3n}{4} \tag{24}$$

Hence, ARE(U,q') is greater than 1 for all n.

## 6. Conclusion

It has been shown that by a suitable extension of the double median test for trend, a more efficient test results, relative to the original test. This more efficient test has been referred to as Generalized Blomqvist Correlation of order n, where n is the number of regions into which the x and y axes of the xy plane have been divided by n-1 order statistics from each sample.

There is a tradeoff, though, in using GBC. Although GBC is more efficient in a statistical sense, it involves more computation than does the double median test. Both tests have been used as part of a computer program for analyzing images. The task of computing either correlation coefficient was divided into two parts. First, the sample order statistics were computed, and in the second part, the N samples were classified into the appropriate regions of the xy, and the coefficient of correlation was computed.

The algorithm used for the first part can be found in Aho et al. [9]. Figure 2 gives a comparison of the asymptotic time bounds for the execution of each half of the computation of the correlation coefficient. All logarithms are to the base 2. In computing q', one can easily see that this is simply computing the statistic U when n=2. Hence the computational complexity increases linearly with n.

Further work in this area might be concerned with determining if GBC is more efficient than similar parametric tests. This question has not been addressed in this paper. Another generalization investigated could be similar to that of Mosteller [5] in letting the order statistics move and choosing those order statistics that give optimal results for some appropriate criteria. Also not addressed was the case where a different number of statistics from the y sample is chosen, say m, so that the division of the xy plane is not into $n^2$ regions but nm where $n \neq m$. This, and other interesting questions concerning distribution-free statistics, remain to be answered.

## References


1. Kruskal, W. H., "Ordinal Measures of Association," Journal of the American Statistical Association, vol. 53, 284 (December 1958), pp. 814-861.

2. Blomqvist, Nils, "On a Measure of Dependence Between Two Random Variables," Annals of Mathematical Statistics, vol. 21, 4 (December 1950), pp. 593-600.

3. Elandt, Regina C., "Exact and Approximate Power Function of the Nonparametric Test of Tendency," Annals of Mathematical Statistics, vol. 33, 2 (June 1962), pp. 471-481.

4. Gibbons, J. D., Nonparametric Statistical Inference, McGraw-Hill, New York, 1971.

5. Mosteller, F., "On Some Useful Inefficient Statistics," Annals of Mathematical Statistics, vol. 17, 4 (December 1946), pp. 377-408.

6. Bradley, J. V., Distribution-Free Statistical Tests, Prentice-Hall, Englewood Cliffs, New Jersey, 1968.

7. Hoeffding, W., "A Class of Statistics with Asymptotically Normal Distribution," Annals of Mathematical Statistics, vol. 19, 3 (September 1948), pp. 293-325.

8. Randles, R. H., and D. A. Wolfe, Introduction to the Theory of Nonparametric Statistics, John Wiley & Sons, New York, 1979.

9. Aho, A. V., J. E. Hopcroft, and J. D. Ullman, The Design and Analysis of Computer Algorithms, Addison-Wesley, New York, 1974.

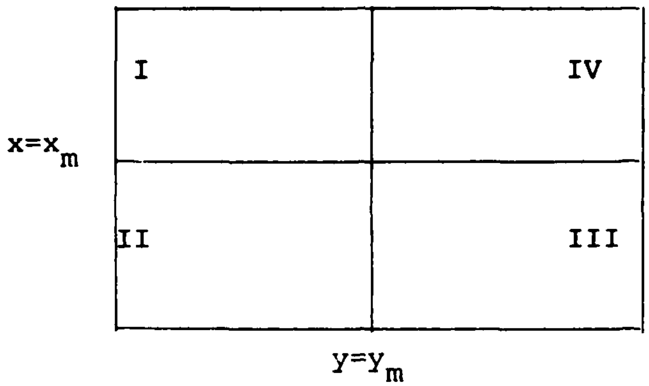


Figure 1. xy plane divided by the lines $x=x_m$ and $y=y_m$.

| | part 1 | part 2 |
|---|---|---|
| q' | 2 logN | N |
| U | 2(n-1)logN | Nlogn |

Figure 2. Asymptotic time bounds for computing q' or U.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

| REPORT DOCUMENTATION PAGE | | READ INSTRUCTIONS BEFORE COMPLETING FORM |
|---|---|---|
| 1. REPORT NUMBER<br>AFOSR-TR- 81 -0560 | 2. GOVT ACCESSION NO.<br>AD-A102 058 | 3. RECIPIENT'S CATALOG NUMBER |
| 4. TITLE (and Subtitle)<br>GENERALIZED BLOMQVIST CORRELATION | | 5. TYPE OF REPORT & PERIOD COVERED<br>INTERIM |
| | | 6. PERFORMING ORG. REPORT NUMBER<br>TR-1043 |
| 7. AUTHOR(s)<br>Stanley M. Dunn | | 8. CONTRACT OR GRANT NUMBER(s)<br>AFOSR-77-3271 |
| 9. PERFORMING ORGANIZATION NAME AND ADDRESS<br>Computer Vision Laboratory<br>Computer Science Center College Park,<br>University of Maryland MD 20742 | | 10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS<br>61102F<br>2304/A2 |
| 11. CONTROLLING OFFICE NAME AND ADDRESS<br>Math & Info. Sciences, AFOSR/NM<br>Bolling AFB<br>Washington, DC 20332 | | 12. REPORT DATE<br>April 1981 |
| | | 13. NUMBER OF PAGES<br>18 |
| 14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) | | 15. SECURITY CLASS. (of this report)<br>Unclassified |
| | | 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE |

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Random variables
Dependence
Distribution-free
Quadrant measure
U-statistics

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

A distribution-free measure of dependence between two random variables is defined and shown to be a generalization of the quadrant measure of association. This new measure is presented as a U-statistic and its asymptotic efficiency relative to the quadrant measure is given. This measure can be used for locating edges in computer images.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

DATE
ILMED
— 8